GREGOR BOEHL
THOMAS FISCHER

Can Taxation Predict US Top-Wealth Share Dynamics?

Institute for Monetary and Financial Stability
GOETHE UNIVERSITY FRANKFURT

WORKING PAPER SERIES NO. 118 (2017)

Institute for Monetary and Financial Stability
Goethe University Frankfurt
House of Finance
Theodor-W.-Adorno-Platz 3
D-60629 Frankfurt am Main
www.imfs-frankfurt.de | info@imfs-frankfurt.de

Can Taxation Predict US Top-Wealth Share Dynamics?

Gregor Boehl* and Thomas Fischer†

December, 2017

Abstract

Yes, the level of capital gains taxation has high explanatory power. We develop a micro-founded portfolio-choice model where idiosyncratic return risk and disagreement in expectations on asset returns generate an analytically tractable fat-tailed Pareto distribution for the top-wealthy. Wealth concentration is dampened by the degree of capital gains taxation. The model is estimated using Kalman filtering and provides good out-of-sample forecasts for both levels and dynamics of wealth concentration in the USA. We show that the tax rate explains historical trends in wealth inequality precisely, and make predictions about the future evolution.

Keywords: Wealth inequality, US top-wealth shares, capital taxation, Fokker-Planck equation, Kalman Filter

JEL No. D31, H23, G11

*Institute for Monetary and Financial Stability, Goethe University Frankfurt, Germany. Email: boehl@econ.uni-frankfurt.de

†Department of Economics and Knut Wicksell Centre for Financial Studies, Lund University, Sweden. Email: thomas.fischer@nek.lu.se

[1]We are grateful to Christiane Clemens, Alex Clymo, Herbert Dawid, Cees Diks, David Edgerton, Frederik Lundtofte, Ben Moll, Salvatore Morelli, Christian Stoltenberg and participants of several conferences and workshops for discussions and helpful comments on the contents of this paper.

One of the oldest debates in economic science is the fundamental question of what drives economic inequality. Some economists (e.g. Piketty, 2014) see an inherent higher order at play, suggesting that a trend of increasing wealth concentration is an *inbuilt* property of market economies. The answer we propose in this paper is far more mundane. While in a *laissez-faire* economy wealth inequality would indeed explode, the distribution of wealth in modern economies is shaped by the structure of the tax system and in particular by the taxes on capital gains. As the structure of the tax system is determined by the policy makers, they can also shape the distribution of wealth. This is of specific relevance in the light of what US President Donald Trump calls *"massive tax cuts"*.

The topic of inequality lay dormant for a long time until recently the French economist Thomas Piketty put the issue back at the forefront of the academic debate by documenting that after a period of contraction, the concentration of economic resources again increased in the 1980s. This is notably the case for the stock measure of wealth. Piketty (2014) focuses on a small group of wealth holders – the top percentiles. For these individuals, the main source of income is capital income as opposed to labor income, which also implies a high inequality of the flow measure income. While offering a detailed description of the issues, little formal explanation for the observed behavior is presented and his future outlook is of rather speculative nature.

This paper attempts to fill this research gap. We develop a micro-founded model that is able to forecast the evolution of the top shares of the wealth distribution. We estimate the model for the US case, which has witnessed a remarkable increase in wealth inequality. The exogenous variable driving the evolution of wealth inequality dynamics – as suggested by Piketty (2014) – is the taxation of capital gains. Without redistributive taxation, we get explosive wealth inequality while higher taxation lowers steady state inequality. In line with the recent empirical literature we focus on the top tails starting from 1954. This extents the perspective as compared to other recent papers trying to match the US evidence, which start from the 1970s (Hubmer et al., 2016; Aoki and Nirei, 2017).

Given the accuracy of the estimated model in terms of out-of-sample forecasts, we also make forecasts for alternative tax regimes. Remaining at the level of taxation initiated by the government under President Obama would in fact considerably decrease the degree of wealth concentration, whereas (*"massive"*) tax cuts back to the pre-Obama level would further increase wealth inequality in the USA. In the latter case, concentration has not yet converged to its new steady state.

Formally, we employ a model of the *random-growth* class which is analytically tractable and shown to exhibit Pareto tails under fairly general conditions (cf. Benhabib et al., 2011). This type of model usually exhibits transition dynamics that are too slow to match the empirical evidence (Gabaix et al., 2016). The previous literature in this field emphasizes the effect of multiplicative idiosyncratic capital income risk as a driver of inequality, as opposed to labor income risk, which is additive. We supplement our model with disagreement about the future returns of financial assets as an additional mechanism to drive the dispersion of wealth, which overcomes the problem of slow convergence dynamics and enables us to fit the dynamics. Since we center our analysis around equity trading which is a central part of wealthy-individual portfolios (Saez and Zucman, 2016), we do not include characteristics which are of importance for the left tail of the distribution (the *poor*).[1]

Going beyond pure numerical simulation, we are able to not only quantify the stationary distribution but the whole dynamics of the top wealth shares in a closed form manner. Disagreement, i.e. heterogeneity about future asset prospects, increases inequality while higher idiosyncratic risk lowers wealth inequality since it is internalized in agents' portfolio decisions. In terms of dynamics, both higher taxes and a larger disagreement increase the convergence dynamics. From a policy perspective, this implies that the reduction of inequality after a tax hike is faster than the increase following a tax cut of the same magnitude.

The remainder of this work is structured as follows. In Section 1 we provide an overview of the empirical and theoretical literature on wealth inequality with a focus on recent papers that attempt to fit the empirical evidence. In the following section we present the micro-foundations for our formal model and discuss analytically statistical properties in Section 3. Section 4 uses the model to generate forecasts about the future evolution of wealth inequality and presents robustness checks. Finally, Section 5 concludes the paper.

1 Literature

Following the major public debate surrounding the publication of the work of Piketty (2014), the empirical evidence regarding inequality – especially for the flow measure of income – has substantially improved. Cross-country evidence is assembled and made

[1] A non-exhaustive list features entry barriers into financial markets in favor of the wealthy, portfolio return and risk increasing with wealth (which can be e.g. captured by a utility function with decreasing relative risk aversion), and inter-generational wealth transmission.

freely available on the *World Income & Wealth Database* maintained by the collaborative effort of many researchers. Despite this effort, the data availability of consistent and long-run measures of wealth inequality is still highly limited. The database provides long-run data for the United States of America, France and the United Kingdom. The US data – important for our paper – was recently updated by Saez and Zucman (2016). [2] A recent comprehensive survey on the overall empirical evidence regarding wealth inequality is given in Roine and Waldenström (2015). The discussion about the distribution of wealth is not only an end in itself, but also contains important policy implications as it impacts on the conduct of monetary policy (Kaplan et al., 2016), on economic growth (e.g. Clemens and Heinemann, 2015) and on financial stability (Kumhof et al., 2015).

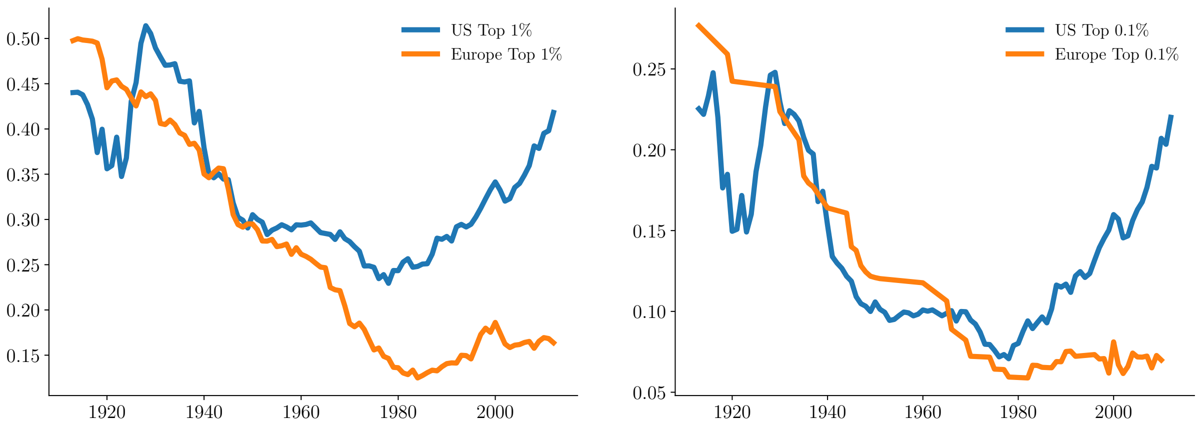

Figure 1: Top wealth shares for the USA and selected European countries

Notes: Wealth inequality decreased until the 1980s. While it leveled in Europe, it increased in the USA afterwards. Left panel includes the UK, France and Sweden, right panel only France and Sweden. Missing values are interpolated. Data source: wid.world and Lundberg and Waldenström (2017) for Sweden.

Figure 1 presents collected evidence on the top-shares for the USA and the European Countries – France, Sweden and the UK – in the long run. While it displays an overall decrease in wealth concentration in all countries until the 1980s, inequality has since increased. This increase is modest in the European countries, but highly pronounced in the USA and in particular emerges for the top wealth holders.

Different theoretical models compete in order to explain the observed degree of inequality. Usually, models in the Bewley-type tradition are considered in order to discuss

[2] The data is available at wid.world. For the UK, the latest data update was conducted by Alvaredo et al. (2017). The quality of the French data (especially from the 1970s onward) was recently substantially improved by Garbinti et al. (2017). Evidence for Sweden is compiled by Daniel Waldenström and his collaborators (Lundberg and Waldenström, 2017) and is freely available on his homepage.

inequality (Bewley, 1986; Huggett, 1993; Aiyagari, 1994). Yet, it has been formally shown by Benhabib et al. (2011) that these types of models – built around the notion of additive idiosyncratic labor income risk – will fail to generate the fat tails in the wealth distribution and thus match the shares of the top wealth holders. Benhabib et al. (2011) propose a model with multiplicative idiosyncratic capital income risk in order to replicate the current state of wealth inequality in the USA. They follow an argument laid out as early as Wold and Whittle (1957), building on random growth (hence the term *random growth models*). Taxation of capital (income) plays a crucial role in these models. Using simulations, Fernholz and Fernholz (2014) show that wealth inequality does explode without redistribution in a standard model with idiosyncratic investment risk. Thus, a tax that addresses these multiplicative shocks on capital returns – i.e. the capital gain tax – is also crucial to understand top dynamics. As documented in Saez and Zucman (2016), the saving rate increases for the top wealth holders. Accordingly, the prime share of their income originates from saved wealth rather than labor income (also cf. Piketty, 2014).[3]

While heterogeneous portfolios are often motivated by different degrees of risk aversion and marginal propensities to consume, this argument does not hold for the very rich since they should have a relatively similar portfolio structure. Even if the portfolio structure would be alike, given the large supply of similar assets within classes this would not imply that portfolios are identical. But due to heterogeneous individual expectations about future prospects, agents still hold different positions among asset classes. For that reason we explicitly motivate heterogeneous portfolios by marginal disagreement on future returns, of which a considerable degree is documented by Greenwood and Shleifer (2014) in a survey over six data sets on investor expectations of future stock market returns. Pfajfar and Santoro (2008, 2010) provide empirical evidence in support of heterogeneous expectations using survey data on inflation expectations. Additional evidence from the lab has shown that individuals generally do not perform well when forming expectations and these expectations are furthermore largely heterogeneous, which is summarized in e.g. Hommes (2013). Boehl (2017) shows that expectations will be heterogeneous even if a considerable fraction of traders is *super-rational*. Recently the heterogeneous expectations hypothesis has made its way into macroeconomics (Mankiw et al., 2003; Branch, 2004).

Given the new data evidence, similar projects have been undertaken. Most prominently, Kaymak and Poschke (2016) use the evidence for the United States from 1960 to the most present date to present a calibrated model in the Bewley tradition. Using

[3]The reader is especially referred to figures 8.4 and 8.10 in Piketty (2014) for French and US evidence.

the modification of Castaneda et al. (2003) and allowing for extreme *superstar* shocks to reproduce high levels of income inequality, the authors are able to match the data considerably well. Aiming to identify the contributing factors with a very detailed modeling of the US-tax system (including income, corporate, and estate taxes as well as the pension system), the authors argue that the structure of the taxation and transfer system is key to explaining the evolution of wealth inequality. Hubmer et al. (2016) extend an otherwise standard Bewley-type model with heterogeneous rates of time preference (Krusell and Smith, 1998), Pareto tails in the income distribution and idiosyncratic investment risk (Benhabib et al., 2011). They reproduce wealth inequality dynamics as a result of substantial tax changes, with a data scope starting from the 1970s. While concentration has started to increase in this period, this omits the relatively stable period of the 1960s and the period of decreasing wealth concentration in the 1970s.[4] They conclude that the substantial increase in income inequality, the change in labor share, the gap between the interest rate and the growth rate $r > g$ (Piketty, 2014) all fall short of explaining these dynamics.[5]

A different approach is presented in Aoki and Nirei (2017), featuring a rich model in continuous time. In line with empirical evidence and due to idiosyncratic firm shocks the emerging stationary distribution of firms is given by Zipf's law.[6] The firm's returns translate into income for private households, implying a realistic distribution of both income and wealth for private households. Combining this framework with a set of tax rates they are able to match both the dynamics and the state of inequality in the USA. Again, the authors focus on the data from the 1970s to the most recent years. Cao and Luo (2017) introduce idiosyncratic return risk into an otherwise standard neoclassical growth model to account for US wealth inequality and also show that the latter is accompanied by increasing capital-output-ratios and decreasing labor shares. In contrast to their work, our paper focusses on the distributional impact on taxation and does not make a statement about the macroeconomic impact of the wealth tax or even its optimal level.[7]

[4]Although *ex-ante* heterogeneous agents with heterogeneous time preference rates are a technical vehicle that can replicate wealth inequality, we would argue that it does not entirely convince as the empirical driver of concentration processes.

[5]In the core, Piketty argues that wealth inequality will not converge as long as the rate of interest r is larger than growth rate of labor income g. A detailed formal and critical discussion of this argument is presented in Fischer (2017).

[6]The latter is a Power-law with an exponent $\alpha = 1$.

[7]Judd (1985) shows that in standard models the optimal tax on the stock value of wealth is zero. It is well-known that in Bewley-type models this result fails to hold and optimal taxes need to be positive in order to counteract excessive savings (Aiyagari, 1995). More recent contributions discussing the welfare impact in both the state and the transition of wealth taxes in the broader sense are e.g. Castaneda et al.

2 Model

We assume an economy with a very large number of individuals indexed by i. In line with the standard literature we assume that time is discrete.[8] Their only income consists of investment returns and they are free to choose between a risk-free asset paying a constant gross return of R and a continuum of ex-ante identical risky assets of which each pays an idiosyncratic, stochastic dividend $d_{i,t}$ every period t. To maximize their intertemporal consumption over an infinite time horizon the agents accumulate wealth $w_{i,t}$. Hence, each agent i faces the question of which amount $c_{i,t}$ to consume and which amount $x_{i,t} = z_{i,t}w_{i,t}$ of the risky asset to purchase. In this case $z_{i,t}$ relates the demand for risky assets as a share of individual wealth $w_{i,t}$. As pointed out earlier, such a model provides a realistic representation for the behavior of wealthy agents, but – in the absence of features such as borrowing constraints and labor income – naturally falls short in the context of the lower 50% share of wealth holders. Since we aim to explain the dynamics of top-shares such simplification is justifiable because, trivially, the wealth share of the bottom $(1-x)$% can be seen as the residual wealth not owned by the top x%.

Assuming log-preferences, the individual problem is then given by

$$\max_{c,z} E_t \sum_{t=0}^{\infty} \beta_t^t \ln c_{i,t}$$

subject to the two constraints

$$w_{i,t} = \left(R + [d_{i,t} + p_t - Rp_{t-1}]z_{i,t-1}\right) s_{i,t-1}(1-\tau)w_{i,t-1}, \tag{1}$$

$$c_{i,t} = (1 - s_{i,t})w_{i,t}. \tag{2}$$

Here we denote by $\beta_t = \beta + \epsilon_t^\beta$ the intertemporal discount rate with an iid. zero-mean preference shock and by $s_{i,t}$ the savings rate. p_t is the price for an asset of the class of risky assets in t and $d_{i,t} = d + \epsilon_{i,t}^d$ its dividend with an idiosyncratic stochastic term $\epsilon_{i,t}^d \sim N(0, \sigma_t^d)$ where $\sigma_t^d = \sigma_d + \epsilon_t^d$ is again subject to an iid. time-varying aggregate shock.[9] The value τ captures a tax on the stock level of wealth.[10] We want to assume that our

(2003), Domeij and Heathcote (2004), and Cagetti and Nardi (2009).

[8]Note that in order to find the cross-sectional distribution using the Fokker-Planck equation (cf. Section 3) we have to transfer to a continuous time approach. Given the discrete nature of data, in our empirical application (Section 4) we afterwards return to a discrete time setting.

[9]Since those assets are ex-ante identical, their price is likewise the same.

[10]Note that in the empirical part it is very important that taxes vary in time. For the sake of readability

taxation is a redistributive transfer towards the bottom shares of society. Given that we model the shares of the top wealthy, any positive lump-sum transfers are negligible. The above problem does not directly entail a closed form solution, but can be separated into two stages that both are relatively standard in the literature. Let us first solve the consumption problem.

Levhari and Srinivasan (1969) show that for log-utility, which is a particular case of Constant Relative Risk Aversion (CRRA) preferences, in equilibrium agents consume $1 - \beta_t$ of their wealth at the end of each period, i.e. $s_{i,t} = \beta_t \; \forall i, t$.[11] It is important to point out that this result holds despite the tax rate. Due to the exact offsetting of income and substitution effects for log-utility the savings rate is not distorted by the tax rate.[12] Note that the assumption of CRRA also explicitly avoids inequality dynamics induced by a different marginal propensity to consume. Thus, the law of motion for each individual's wealth follows

$$w_{i,t} = (1-\tau)\beta_t R^z_{i,t}(z_{i,t-1})w_{t-1}, \tag{3}$$

for which $R^z_{i,t}(z_{i,t-1})$ summarizes the individuals gross return on investment.

For the second stage, in which we solve for the optimal demand for risky asset $x_{i,t}$, let us use Equation (1) to rewrite the maximization problem as

$$\max_z E_t \sum_{k=0}^{\infty} {\beta_{t+k}}^{t+k} \ln\{(1-\beta_{t+k} w_{i,t+k}\} \quad \text{s.t. } w_{i,t} = (1-\tau)\beta_t R^z_{i,t}(z_{i,t-1})w_{i,t-1} \forall t \in \mathbb{N}$$

which is equivalent to

$$\max_z E_t \sum_{k=0}^{\infty} {\beta_{t+k}}^{t+k} \ln\left\{(1-\beta_{t+k})\beta_{t+k} w_{i,t}(1-\tau)^k \prod_{l=0}^{k} R^z_{i,t+l}(z_{i,t+l-1})\right\}.$$

Due to the logarithmic laws the term $\ln\{\prod_{l=0}^k R^z_{i,t+l}(z_{t+l-1})\}$ can be separated and is the only part that depends on z_t. Since we can rewrite $\prod_{l=0}^k R^z_{i,t+l}(z_{i,t+l-1}) = R^z_{i,t}(z_{i,t-1}) \prod_{l=1}^k R^z_{i,t+l}(z_{i,t+l-1})$

we, however, suppress the time index in this section.

[11] Levhari and Srinivasan (1969) derive a more general result for CRRA utility ($u(c) = \frac{c^{1-\gamma}}{1-\gamma}$) and iid. returns. Depending on whether income ($\gamma > 1$) or substitution effects ($0 < \gamma < 1$) dominate, individuals adjust their savings taking into account the risky savings technology. The special case of perfectly offsetting income and substitution effects ($\gamma = 1$) assumed here, implies that the nature of the stochastic returns has no impact on the savings decision.

[12] The interested reader is also referred to Lansing (1999), who shows that the seminal result of a zero optimal tax rate as proposed in Judd (1985) fails to hold with log-utility. For a more recent and general approach the reader is referred to Straub and Werning (2014).

this portfolio problem can be well-approximated by mean-variance maximization, as laid out in Pulley (1983). The optimal demand for the risky asset $x_{i,t}$ is then, up to a second order approximation, given by

$$x_{i,t} = z_{i,t} w_{i,t} = \left(E_t[d_{t+1} + p_{t+1}] - Rp_t\right) w_{i,t} / \sigma_t^{d^2}. \tag{4}$$

Note that – identical to the optimal consumption plan – the portfolio structure is independent of the wealth tax. As presented in Stiglitz (1969) for Constant Relative Risk Aversion preferences – of which the assumed log-utility is a special case – wealth taxation does not lead to a restructuring of the portfolio.

Let us assume that return expectations are heterogeneous and each agent's expectation is a draw from the normal distribution around the rational expectation of future returns. Thus, the rational expectation operator E is replaced with a noisy individual expectation operator $\hat{E}_{i,t}$, giving

$$\hat{E}_{i,t}[d_{t+1} + p_{t+1}] = d + E_t p_{t+1} + \epsilon_{i,t}^E, \qquad \epsilon_{i,t}^E \sim N(0, \sigma_t^E),$$

where $\sigma_t^E = \sigma_E + \epsilon_t^\sigma$ as well can be subject to the iid. time-varying aggregated news shock ϵ_t^σ.

Assume furthermore that no single person is rich enough or has an $\epsilon_{i,t}^E$ large enough to influence the price.[13] Market clearing requires $\sum_i x_{i,t} = X_t$, with X_t being the total supply of the risky asset. Without loss of generality we can fix supply and normalize $X_t = 1$ to unity for all periods.

Keeping this in mind and aggregating over Equation (1) and (4) yields

$$p_t = R^{-1}(E_t p_{t+1} + d_{t+1} - \sigma_t^{d^2} W_t^{-1}) \tag{5}$$

$$W_t = \beta(RW_{t-1} + d_t + p_t - Rp_{t-1}), \tag{6}$$

which is the law of motion for prices and aggregated wealth W_t. Note that for the stationarity of aggregate wealth redistribution of tax proceedings is required. If this were not the case, in the long run all private wealth would be transferred to the government. If we assume that all variables are detrend, due to the law of large numbers idiosyncratic disturbances level out and aggregate wealth $W_t = W$ is constant in the absence of aggregate

[13] This is indeed satisfied by the law of large numbers. The additional advantage of this assumption is, without loss of generality, that we can provide analytic results for the law of motion of individual wealth, aggregated wealth, and prices.

shocks. The values of prices and aggregated wealth thus reflect the detrend steady growth path.[14] Then, we can also normalize the price to unity without explicitly accounting for market clearing. The steady state versions of (5) and (6) are

$$\sigma_d^2/W = d + 1 - R \tag{7}$$

$$W(\beta^{-1} - R) = d + 1 - R. \tag{8}$$

This implies that, given the normalization of prices,

$$W = \frac{\sigma_d}{\sqrt{\beta^{-1} - R}} \tag{9}$$

$$d + 1 - R = \sqrt{\beta^{-1} - R}\sigma_d. \tag{10}$$

Plugging Equation (4) into Equation (1), integrating individual forecast errors and setting prices to the steady state yields

$$w_{i,t} = \beta_t \left\{ R + \left(d + 1 + \epsilon_{i,t}^d - R\right)\left(d + \epsilon_{i,t}^E + 1 - R\right)\sigma_d^{-2} \right\}(1 - \tau)w_{i,t-1}. \tag{11}$$

which together with Equation (10) and some algebra can be written as the law of motion (LOM) for individual wealth

$$w_{i,t} = \beta_t \left\{ \beta_t^{-1} + \sqrt{\beta_t^{-1} - R}(\epsilon_{i,t}^d + \epsilon_{i,t}^E)/\sigma_t^d + \epsilon_{i,t}^d \epsilon_{i,t}^E {\sigma_t^d}^{-2} \right\}(1 - \tau)w_{i,t-1}.$$

We use annual data, so let $\beta = 0.95$. For realistic values of the mean real interest rate R we have $\sqrt{\beta^{-1} - R} \approx 0$ to be negligibly small.[15] After defining $\gamma_t \equiv \beta_t \frac{\sigma_t^E}{\sigma_t^d}$ and $\varepsilon_{i,t} \equiv \epsilon_{i,t}^d \epsilon_{i,t}^E$ to be the product of two independent random variables that follow a standard normal distribution, the final law of motion can be further simplified to

$$w_{i,t} = (1 + \gamma_t \varepsilon_{i,t})(1 - \tau_t)w_{i,t-1},$$

γ_t now contains the aggregate shocks ϵ_t^β, ϵ_t^d and ϵ_t^σ and the expected value γ remains as

[14] This assumption implies that all growth in aggregate wealth can be attributed to some uniform exogenous growth rate. Note that the latter does not distort distributional properties.

[15] Note that a positive demand for the risky assets requires $R < 1 + d$. Stationarity of aggregate wealth furthermore demands for $R < \beta^{-1} < 1 + d$ i.e., $\sqrt{\beta^{-1} - R} > 0$ but small. Moreover, the variance of $\epsilon_{i,t}^E$ is already quite minor. Rewriting $\epsilon_{i,t}^d$ in terms of a standard normal reveals that the term is relatively small.

the only free parameter of our model.

3 Representation in Closed Form

This section aims to enrich our understanding of the process that generates the wealth distribution by finding a closed form solution for the stationary distribution as well as for the transition dynamics. In order to do so, we have to overcome some technical obstacles. For better readability we will omit the aggregate shocks until the end of the section as they do not have an impact on the shape of the distribution.

3.1 Cross-sectional distribution

The portfolio returns, a product of two standard normal variables, follow a so-called *product-normal distribution*. To obtain a closed form solution, we have to transfer this distribution to another distribution that can be handled analytically.

Proposition 1. The first three moments of the product normal distribution and the Laplace distribution with shape parameter of $\lambda = \sqrt{0.5}$ are equal.

Proof. See Appendix A. ■

The Laplace distribution is very handy in our context for identifying a closed-form solution. The individual law of motion (LOM) has to be rewritten in continuous time in order to solve the Fokker-Planck equations which allows us to identify the cross-sectional distribution in terms of the free parameters γ and τ. It then reads as

$$dw_i = -\tau w_i dt + (1-\tau)\gamma w_i dNP, \tag{12}$$

for which NP is the noise following the product-normal distribution. In order to retrieve a closed-form solution we transform this to the Laplace distribution using the scaling factor λ, which we just introduced. The equation thus reads

$$dw_i = -\frac{\tau}{\lambda} w_i dt + \frac{1}{\lambda}(1-\tau)\gamma w_i dL, \tag{13}$$

for which L signifies Laplace distributed noise.

Proposition 2. Using Itô's lemma as a second-order approximation and solving the Fokker-Planck equation, the right tails of the cross-sectional distribution (the top wealth holders) are described by a Pareto distribution with a tail parameter α,

$$\alpha = 1 + \frac{\sqrt{2}\tau}{\gamma^2(1-\tau)^2}. \tag{14}$$

Proof. Let us define the log of wealth $\hat{w}_{i,t} = \log(w_{i,t})$ and apply Itô's lemma as a second-order approximation. Thus the equation reads

$$d\hat{w}_i \approx \left(-\frac{\tau}{\lambda} - 0.5\frac{\gamma^2}{\lambda^2}(1-\tau)^2\right)dt + \frac{1}{\lambda}(1-\tau)\gamma dL = -\mu dt + \delta dL,$$

with a diffusion term $\delta \equiv \frac{1}{\lambda}(1-\tau)\gamma$ and a drift $\mu \equiv \frac{\tau}{\lambda} + 0.5\frac{\gamma^2}{\lambda^2}(1-\tau)^2 = \frac{\tau}{\lambda} + 0.5\delta^2$. As shown in Toda (2012), the Laplace distribution with unit standard deviation can be modeled by

$$dL = -\lambda\,\text{sign}(L)dt + \ dB$$

with B being the standard Brownian motion and $\text{sign}(x) = \frac{x}{|x|}$ representing the sign function. In essence, this is a Brownian motion which reverts to its zero mean both in the positive and the negative domain. Thus, the noise in the returns ε before taxes (approximately) follows a Laplace distribution with a zero mean

$$f(\varepsilon) = \frac{0.5}{\gamma\lambda}\exp\left(-\frac{|\varepsilon|}{\gamma\lambda}\right),$$

which can be understood as a symmetric double exponential distribution. This result will also be of use in Proposition 7. If we ignore the fat-tail properties in the returns – induced by the mean reversion – we can model the wealth evolution of the wealthiest individuals by

$$d\hat{w}_i = -\mu dt + \delta dB, \qquad \hat{w}_{i,t} >> 0. \tag{15}$$

The cross-sectional distribution can be found by solving the so-called Fokker-Planck equation[16]

$$\frac{\partial f(\hat{w},t)}{\partial t} = -\frac{\partial}{\partial \hat{w}}\left(\mu f(\hat{w},t)\right) + 0.5\frac{\partial^2}{\partial \hat{w}^2}\left(\delta^2 f(\hat{w},t)\right).$$

We first consider the stationary distribution ($\frac{\partial f(\hat{w},t)}{\partial t} \stackrel{!}{=} 0$). The solution is well-known

[16] The latter is frequently referred to as Kolmogorov forward equation. The terms can, however, be used interchangeably.

(Karlin and Taylor, 1981, p. 221) and given by[17]

$$f(\hat{w}) = C\exp(-\alpha\hat{w}), \tag{16}$$

for $\hat{w} > \hat{w}_{min} = \ln(w_{min})$ with an integration constant of $C = w_{min}^{\alpha}\alpha$ to ensure $\int_{\hat{w}_{min}}^{\infty} f(\hat{w})d\hat{w} = 1$. For our case we have

$$\alpha = \frac{2\mu}{\delta^2} = 1 + \frac{\sqrt{2}\tau}{\gamma^2(1-\tau)^2}. \tag{17}$$

It is easy to transfer the exponential distribution to a Pareto distribution. In fact, if $\hat{w}$ follows the described exponential distribution, wealth $w = \exp(\hat{w})$ is given by the probability density function

$$\lim_{w\to\infty} f(w) \sim w^{-\alpha-1}.$$

■

It is important to acknowledge the necessary conditions for this result to emerge. It requires both (i) mean reversion ($\mu > 0$) and (ii) a positive non-zero reflecting barrier ($\hat{w}_{min} > 0$). Note that we do not model the latter explicitly. Yet, one could consider that the overall proceedings of the wealth tax are redistributed to all individuals in an equivalent lump-sum manner. For a given tax rate τ and a stationary average wealth $\bar{w}$ the latter would amount to $\tau\bar{w}$. The two assumptions also have an important economic implication. Mean reversion is achieved by a positive capital tax rate that counteracts the multiplicative stochastic noise of the capital gains. The second condition also ensures that the capital tax is not a net loss for the private households. It moreover ensures overall stationarity of private wealth.

In fact, the complete distribution is characterized by the single value α. Thus, other measures regarding inequality can be derived starting from this assumption.

Proposition 3. The stationary ($t \to \infty$) share $s^x(\tau, \infty)$ of the top x (e.g. the top 1% implying $x = 0.01$) wealth holders is given by

$$s^x(\tau, \infty) = x^{1-1/\alpha}, \tag{18}$$

for which α, as above, is implicitly a function of taxes τ and γ.

[17] A more formal derivation using Laplace-transforms is presented in Appendix B, also determining the average convergence speed.

Proof. The result is well known and can be derived by computing the closed form value of the Lorenz curve given by $L(F) = 1-(1-F)^{1-1/\alpha}$ and then calculating $s^x = 1-L(1-x)$. ■

The same rationale can also be used to derive a closed form expression for the Gini coefficient. In general a high tail coefficient α is accompanied by low inequality.[18] This very neat result has some strong implications for the asymptotic behavior. First of all, without taxation $\tau = 0$ the tail-coefficient is $\alpha = 1$, identical to Zipf's law. In fact, the Gini coefficient then takes the value of $Gini(w) = 1$ and $s_x(\tau) = 1$ for all $x \in (0,1]$, implying total inequality. Thus, in a laissez-faire economy without government intervention, there is no finite level of inequality. In general, inequality increases (α decreases) with γ while decreasing with taxation τ. For the extreme case of $\tau \to 1$ - which can be thought of as a completely egalitarian society - we would have $\alpha \to \infty$, and thus have a Lorenz-curve identical to the 45-degree line and thus no inequality at all.

Note that our proof heavily relies on second-order approximations. This is, however, not problematic for realistic values of $\alpha < 2$, for which only the first two moments exists. [19]

3.2 Convergence dynamics

We can also make a statement about the convergence speed.

Proposition 4. The convergence of the Laplace-transformed pdf ($\mathcal{L}f(\hat{w},t) = F(s,t)$) is given by

$$F(s,t) - F(s,\infty) \sim \exp(-\phi t), \tag{19}$$

with an average convergence speed of

$$\phi = (0.5\gamma(1-\tau)\alpha)^2 . \tag{20}$$

[18]The closed-form value for the Gini coefficient is given by $Gini(w) = \frac{1}{2\alpha-1}$ and decreasing with α for the realistic case of $\alpha > 1$. Note that then it also holds that $\frac{\partial}{\partial\alpha} s^x(\tau,\infty) < 0$.

[19]The Fokker-Planck equation is in turn also a second-order approximation to the more general Master equation while the use of Itô's lemma is also a second-order approximation. For noise generated by a Brownian motion (rather than the product-normal distribution) it would still hold exactly. As a result only degrees of concentration of $s^{0.01} > s^{0.01}(\alpha = 2) \approx 10.6\%$ can be modelled reliably using the closed-form solutions.

For a one period gap it is given by

$$F(s,t) = F(s,t-1)\exp(-\phi) + F(s,\infty)(1-\exp(-\phi)), \tag{21}$$

Proof. See Appendix B. ■

This implies a half-life of $t_{0.5} = \frac{\ln(2)}{\phi}$. In fact the effective taxation τ not only decreases steady-state inequality, but also increases the speed of convergence to the latter. This also means that there is an asymmetry in the convergence. The increase of inequality for low taxes is slower than the decrease after high tax rates. Thus, the positive message for the policy maker is that it is faster to come down to lower inequality rather than to increase the level of inequality. Finally, we use this general result in order to make an approximate statement about the evolution of top-shares which are the focus of the recent empirical literature and thus also take in a central position in this paper.

3.3 Estimation model

Note that so far we ignored the time dimension. Let us assume that the value of γ is constant in time. Yet, due to policy changes the tax rate τ_t is varying in time. Thus, not only the stationary level of inequality as modeled by the Pareto-tail α_t varies in time t, but so does the convergence speed ϕ_t.

Proposition 5. Ignoring aggregated shocks, the top-shares approximately evolve according to an autoregressive process of first-order with

$$s_t^x = \rho_t s_{t-1}^x + (1-\rho_t)s^x(\tau_t,\infty), \tag{22}$$

and $\rho_t = \exp(-\phi_t)$ for the average convergence speed $\phi_t = \phi(\tau_t)$ as defined in Equation (20).

Hence, the share owned by the fraction x of the population is a linear combination of the last period's share and the share of the top x of the stationary distribution given the tax rate τ_t at each time t. Let us now reintroduce the aggregated shocks, which are important for the Kalman filtering. The cross-sectional shocks σ_i^E and σ_i^d – indicated by the subscript i – drive the cross-sectional distribution in the first place and are represented through the transformations above. Aggregated shocks affect *the behavior* of all agents equally and hence can introduce aggregate temporary fluctuations to our law of motion for top shares. We summarize these in a composite shock term ε_t^s.

Proposition 6. As a first-order approximation around the stochastic steady state and using the central limit theorem, the law of motion including aggregate time-varying shocks can be summarized by

$$s_t^x = \rho_t s_{t-1}^x + (1-\rho_t) s^x(\tau_t, \infty) + \varepsilon_t^s \tag{23}$$

$$\varepsilon_t^s \sim \mathcal{N}(0, \sigma_s). \tag{24}$$

Proof. Let us write out that $s_t^x(\gamma(\epsilon_t^\beta, \epsilon_t^d, \epsilon_t^\sigma), \tau_t))$ is a function of the three aggregate unobserved iid. shocks. These shocks operate on the idiosyncratic risk ϵ^d, the standard deviation of disagreement ϵ^σ, and the rate of time preference β. Using the multivariate Taylor approximation around the expected value where all shocks are zero yields

$$s_t^x(\gamma(\epsilon_t^\beta, \epsilon_t^d, \epsilon_t^E), \tau_t) \approx s_t^x(\gamma, \tau_t) + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^\beta}\epsilon_t^\beta + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^d}\epsilon_t^d + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^E}\epsilon_t^E$$

for which

$$\varepsilon_t^s \quad := \quad \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^\beta}\epsilon_t^\beta + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^d}\epsilon_t^d + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^E}\epsilon_t^E$$

is the sum of zero-mean i.i.d. random variables each multiplied by a constant. Applying the central limit theorem this is approximately normally distributed and the result in the proposition follows.

■

3.4 Comparative statics

So far it was assumed that there was a pure substance tax on the stock level of wealth, which is not in place in the US. Yet, the stock level of wealth is subject to other more subtle forms of taxation. In particular, for the case of the USA – our empirical application in the next section – net capital gains are taxed. Of course taxes are only levied on positive measures – i.e. capital gains – and not losses. We thus have to translate between the measures.

Proposition 7. The gross-wealth tax τ given a capital gains tax θ_r can be approximated by finding a τ such that the expected value of after tax returns from a capital gains tax and after tax returns of a gross-wealth tax are equal. Gross-wealth taxes are then given by

$$\tau = \frac{1}{2}\gamma\lambda\theta_r. \tag{25}$$

Proof. Dropping time subscripts for taxes, the after-tax returns given the capital income tax θ_r are

$$\bar{R}_{\theta_r} = 1 + \begin{cases} (1-\theta_r)\gamma\varepsilon_{i,t} & \text{if } \varepsilon_{i,t} > 0 \\ \gamma\varepsilon_{i,t} & \text{if } \varepsilon_{i,t} \leq 0. \end{cases}$$

To use the LOM in Equation (22) we approximate τ given θ_r by finding a τ such that the expected value of $\bar{R}_\tau$ equals the expected value of $\bar{R}_{\theta_r}$. Then, given that $\varepsilon_{i,t}$ approximately follows a Laplace distribution with scale $\lambda = \sqrt{0.5}$, the expected value $E[\varepsilon_{i,t}|\varepsilon_{i,t} \leq 0]$ is the mean of an exponential distribution with inverse scale λ, which is again λ. Then

$$\begin{aligned} E\bar{R}_\tau &= E\bar{R}_{\theta_r} \\ E\left\{(1-\tau)\left(1+\gamma\varepsilon_{i,t}\right)\right\} &= 1 + \gamma P(\varepsilon_{i,t} \leq 0)E[\varepsilon_{i,t}|\varepsilon_{i,t} \leq 0] + (1-\theta_r)\gamma P(\varepsilon_{i,t} > 0)E[\varepsilon_{i,t}|\varepsilon_{i,t} > 0] \\ 1-\tau &= 1 - 0.5\gamma\lambda + 0.5(1-\theta_r)\gamma\lambda \\ \tau &= \frac{1}{2}\theta_r\gamma\lambda, \end{aligned}$$

where $P(\varepsilon_{i,t} > 0)$ denotes the probability that $\varepsilon_{i,t}$ is positive. ■

Finally, our model is fully specified, allowing us to conduct some comparative statics. To obtain some intuition, let us plug Equation (25) into the closed-form solution from Equation (22), and for simplicity take $(1-\tau) \approx 1$. Then

$$s_t^x = \exp(-\phi_t)s_{t-1}^x + (1-\exp(-\phi_t))x^{1-1/\alpha_t},$$

with

$$\begin{aligned} \alpha_t(\theta_{r,t}) \approx& 1 + \frac{0.5\theta_{r,t}}{\gamma}, \\ \phi_t(\theta_{r,t}) \approx& \frac{1}{4}\gamma^2 + \frac{1}{4}\theta_{r,t}\gamma + \frac{1}{16}\theta_{r,t}^2, \end{aligned}$$

for which follows that *if the system is at the steady state*

$$\frac{\partial s_t^x}{\partial \gamma} > 0 > \frac{\partial s_t^x}{\partial \theta_{r,t}} \qquad \text{and} \qquad \frac{\partial|\Delta s_t^x|}{\partial \theta_{r,t}}, \frac{\partial|\Delta s_t^x|}{\partial \gamma} > 0.$$

The weight on the most recent value s_{t-1}^x decreases in the transition speed, which depends positively on taxes $\theta_{r,t}$ and dispersion γ. This means that in terms of inequality dynamics, an increase in dispersion γ is a complement to an increase in taxes and will speed up dynamics. However, in terms of inequality levels these two have opposing effects: an increase in taxes $\theta_{r,t}$ decreases the stationary level of inequality, while a higher value of γ will increase it.

It is also insightful to keep in mind the definition of $\gamma = \beta \frac{\sigma^E}{\sigma^d}$ to decompose the effects. We have a higher degree of wealth inequality (as measured by top shares) for high disagreement σ^E. Meanwhile – and somewhat surprisingly – wealth concentration decreases for high idiosyncratic risk σ^d. The latter is due to the fact that individuals incorporate risk into their portfolio decision by increasing the share of risk-free assets. Finally, the inequality increases with the discount factor β which for the assumed case of log-utility is equal to the savings rate. Thus, high savings are accompanied by higher degrees of wealth inequality. In terms of dynamics, both the higher savings rates and higher expectation disagreement increases the dynamics, whereas higher idiosyncratic risk slows down the dynamics.

4 Kalman Filtering

In this section we use the Kalman Filter to estimate the model specified in Equation (23) to the empirical top-wealth shares of the US economy, while feeding-in only the tax data. For robustness, we then perform out-of-sample forecasts. As a last step we forecast the top shares given different scenarios of taxation.

For the top-wealth data we rely on the recent study of Saez and Zucman (2016).[20] The authors incorporate an estimate of offshore wealth when constructing the series in order to capture tax evasion in their top wealth inequality data, which is of particular relevance given our results. The topic is also treated in detail in Alstadsaeter et al. (2017). Note that if tax evasion would be proportional to the level of taxation, this would imply a decrease in the wealth-income ratio when taxes are high, which can not be confirmed by the data from Piketty and Zucman (2014). The data on top-capital gains tax in use is collected by the *Tax Policy Center*.[21] The tax series is then transformed as outlined in Proposition 7. We start our investigation in the year 1954 since from this year on capital gains tax data is available, and end it in 2012, for which the last observation on the top-0.1% is obtainable. This means that the years 2013 and 2014, for which the

[20] The data is freely available on the homepage http://wid.world, featuring long-run time series measuring inequality for several countries.

[21] At the time of writing, the data can be downloaded at
http://www.taxpolicycenter.org/statistics/historical-capital-gains-and-taxes.

US-government under Barack Obama increased the capital gains taxes up to 25.1% is not included in our analysis. We, however, make use of the more recent values of 23.8% (2017) for our predictions further below.

The Kalman filter dates back to Kalman et al. (1960) and is a recursive Bayesian filter that takes into account the measurement errors in the data. Given the model and values of its parameters (for the one dimensional filter this is γ and the exogenous noise σ_s) the filter finds the series of exogenous shocks that is most likely given the data and the standard deviation of exogenous shocks. The input of the tax series is treated as a time-variant parameter. Since the filter returns the likelihood of the data it can be used to estimate the model by choosing parameter values that maximize the likelihood. The respective *filtered* series is then an estimate of the true unobservable series, whereas the estimated standard deviation can be considered as a measure-of-fit for the model. A high standard deviation implies that the model is not a good description of reality, since driving factors are merely exogenous and vice versa. The technical details are relegated to Online Appendix D. Saez and Zucman (2016) indicate a value of measurement errors for the time series of top-wealth between 2% and 3% to be a reasonable estimate. The choice between these values does not have a quantitative impact, for filtering we use the more conservative estimate of 3%.

In the United States, individuals generally pay income tax on the net of their capital gains. There are a considerable number of exemptions, depending on investment duration, net-worth, and general status. The series in use here represents the maximum tax rate on returns with positive net capital gains obtained from the US Tax Foundation. This has the great advantage that it is a tax explicitly and only on capital gains, which are the focus of our model. Yet, we note that reducing a complicated system of tax progression to just one number bears the risk of misalignment. Taxes on labor income and inheritance are ignored since they are beyond the scope of our model.[22]

In Figure 2, we apply the filter to the 1% top shares of the USA. The corresponding estimate of $\gamma = \beta\frac{\sigma^E}{\sigma^D}$ is $\hat{\gamma} = 0.3523$. The finance literature usually uses a value of σ^d in the range from 0.08 to 0.3 (Campbell and Viceira, 2002) on a quarterly basis. Greenwood and Shleifer (2014) estimate a quarterly standard deviation of return disagreement ranging between 1% and 4% . The annual discount rate is assumed 0.95. Combining and annualizing

[22]The latter factor is important for intergenerational mobility and can be captured in a model with Overlapping Generations (OLG). Note that we argue from the perspective of an infinite horizon economy. In fact, the inheritance tax is a specific wealth tax that only becomes relevant at specific points in time (the transition between dynasties marked by the death of the family head). A discussion of the interconnection between demographics and wealth inequality is e.g. given in Benhabib et al. (2014).

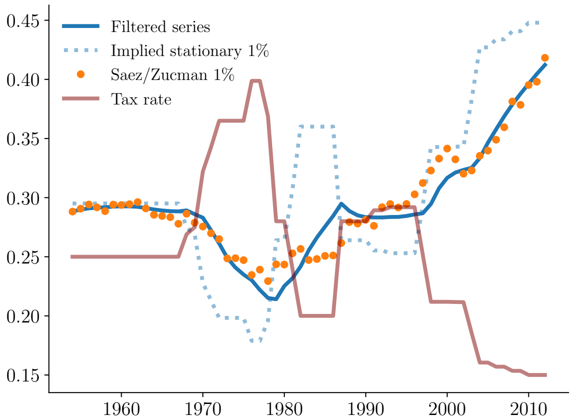

Figure 2: The filtered time series of the top-1% together with the empirical data (dots) and the tax series.
Notes: The dashed line represents the stationary share that is immediately implied by the current tax rate. Input and output data from year 1954 – 2013.

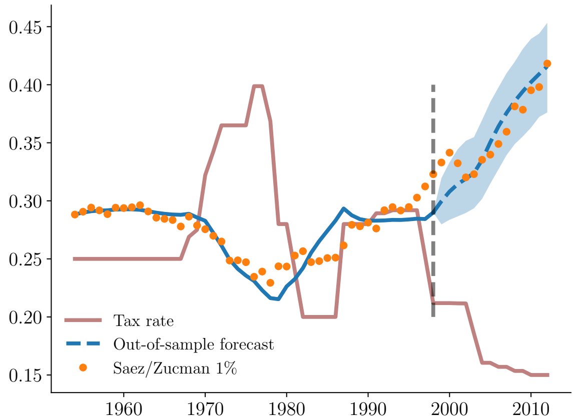

Figure 3: Out-of-sample testing using the time series from years 1954 – 1998.
Notes: Filtered series until year 1998, from then on the dashed line represents the median of simulations together with the 95% prediction intervals.

these values suggests that our estimate for γ lies in the reasonable range from about 0.02 to 0.45 and implies furthermore that the standard deviation of agents' forecasting errors is yet quite small compared to the standard deviations of returns. Thus, the estimated value is in line with relatively low disagreement variance as compared to return variance. The filtered series matches the time series of the data very well, for which the broadly constant share until the early 1970s is in line with a stationary distribution implied by the constant tax rate. The dashed line stands for the analytic result of the stationary cross-sectional distribution implied by the tax rate $\theta_{r,t}$ at time t, i.e. the share to which the distribution will converge if time goes to infinity. For this reason the shares of the stationary distribution *jump* with each change in the tax rate. The filtered series follows more slowly and, in line with the data, slowly converges towards the stationary value. After an increase of capital gains taxes starting in the late 1960s, inequality decreases until the Reagan period, in which taxes return to the previous level. The levels as well as the responses to this tax increase are well matched, while simultaneously providing realistic transition dynamics. The reduction of taxes is accompanied by an immediate increase in wealth inequality both in the data and predicted by our model. While, after a short drop, taxes return to their level from the 1950s in the late 1980s, inequality returns to the same postwar-level in the data as well as in the model. Finally, the tax decreases in the late 1990s and the early 2000s initiate convergence to a steady state of higher wealth inequality that is yet unreached. The estimated standard deviation of the error term, $\hat{\sigma}_s = 0.97\%$, is relatively small given that our model is considerably parsimonious.

To further assess the robustness of the model we conduct out-of-sample forecasts for the top 1% share that are shown in Figure 3. We use the data until the year 1998 as the sample period to estimate γ and σ_s and then run a batch of 1,000 simulations starting in 1998, while again feeding in the respective time series. The year 1998 is chosen because it accommodates an ample decrease in capital gains taxation from 29.19% in 1996 via 25.19% in 1997 towards 21.19% in 1998. While dots and solid line represent the empirical data and the filter respectively, the dashed line depicts the mean of the out-of-sample predictions with the corresponding 95% interval. While the spike in inequality following the year 2000 is unexplained by our model, the data reverts to almost perfectly matching the median convergence path from the mid-2000s onward. Keep in mind that we estimate a long-run mean value of γ and that fluctuation around this mean are integrated in the shock term. The early 2000s were shaped by the dotcom bubble. In the logic of our model, this is captured by a temporary higher disagreement σ_t^E, implying a larger γ_t and thus a higher wealth inequality.

In Online Appendix E we discuss in detail how much our results depend on the selection of the sample period. To summarize, while the *last* periods in the data are not crucial for a good fit, the 50s and 60s are important due to the relative stationarity during this period. This is of notable importance since these periods are omitted in the work of others discussed earlier (Hubmer et al., 2016; Aoki and Nirei, 2017).

In the following we want to incorporate multiple time series and thus employ a multi-dimensional filter. We redirect the treatment of the 2-D filter to Online Appendix F and turn to the 3-D filter right away. Although the upper tail of wealth is often approximated to follow a Pareto distribution, as discussed in Blanchet et al. (2017), this does not match the actual distribution precisely. As presented in Figure 5 the local Pareto coefficient is larger the more we go into the tails of the distribution i.e. for lower values of x.[23] A true Pareto distribution however is scale-free and thus exhibits the identical Pareto coefficient regardless of the level of x. To account for different Pareto coefficients in the

[23] We compute the local Pareto coefficient using the equation $s^x = x^{1-1/\alpha} \leftrightarrow \alpha = 1/(1 - \ln(s^x - x))$. Similar evidence, is e.g. reported in Saez and Stantcheva (2016). They show that the local Pareto coefficient of capital income increases in the tails. Even many years after the initial claim of Pareto (1896) there is still disagreement about the notion of whether wealth follows a true Pareto distribution (Clauset et al., 2009; Chan et al., 2017; Vermeulen, 2016). To summarize the debate briefly, the Pareto distribution is a good approximation which has some pleasant analytic properties, yet fails to match the data precisely with the limitations pointed out above.

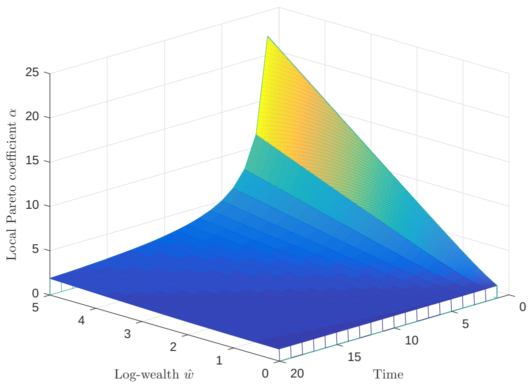

Figure 4: Local Pareto coefficient for log-wealth $\hat{w}$ in time for the model.
Notes: During transition the coefficient increases for the top shares, after convergence the coefficient is independent of the level of wealth.

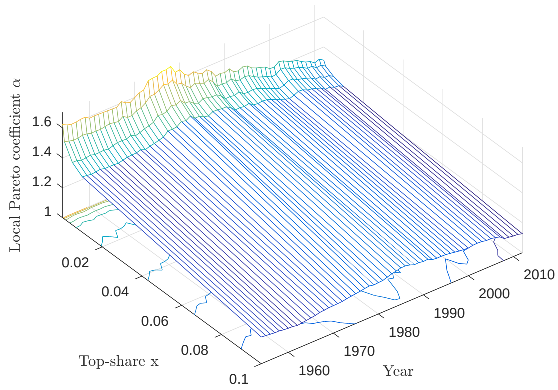

Figure 5: Local Pareto coefficient for top-shares x in time for the US evidence.
Notes: The coefficient does not only vary across time, but also persistently increases towards the tail.

data our estimation problem is adjusted slightly by allowing to estimate a specific γ^x for each different percentile. This also adjusts for the fact that in Proposition 3 the standard representation of Pareto shares is used, which only holds as an approximation. A summary of the estimated parameters can be found in Table 1. As expected, the estimates for each share are almost independent of the number of dimensions of the Kalman Filter.

Table 1: Parameter estimates for the filter using different series of top-shares as input.

Top-shares	$\hat{\gamma}_{1\%}$	$\hat{\gamma}_{0.1\%}$	$\hat{\gamma}_{0.01\%}$	$\hat{\sigma}_s$
1	0.3523	–	–	0.97%
1, 0.1	0.3522	0.3160	–	0.85%
1, 0.1, 0.01	0.3695	0.3291	0.3005	0.73%

Notes: The estimated standard errors are also a measure of the goodness of fit. In line with Figure 5 the value of γ decreases for higher quantiles.

In order to match more extreme tails, the estimated value of γ decreases (cf. Table 1). Put differently, the local Pareto coefficient $\alpha(x)$ increases for more narrow top-shares x. These differences in γ^x across shares can be explained well by higher variances in portfolios, i.e. the very rich have riskier portfolios than the "plain" rich as captured by higher values of σ_d and thus lower values of γ. In Appendix C we discuss the (non-stationary) closed-form solution of the Fokker-Planck equation. It turns out that in the short run the distribution resembles a log-normal distribution with increasing Pareto coefficients in the tails.[24] The increase in inequality is slowly transmitted to the fat ends of the tails. In line with Gabaix et al. (2016), in this type of model the convergence is slower in the tails. In fact, in the short run for the non-stationary distribution the local Pareto coefficient increases in the tail for high values of log-wealth $\hat{w}$ (cf. Figure 4).[25] This is in line with the empirical results as presented in Figure 5.

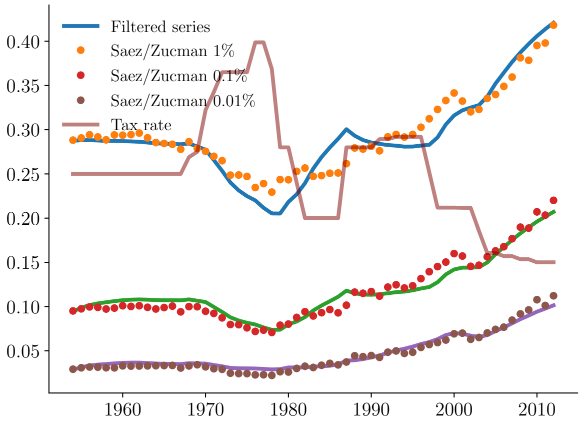

Figure 6: The filtered time series of the top-shares together with the empirical data (dots) and the tax series for three different top-shares.
Notes: The fit is created using the 3-dimensional Kalman Filter. Input and output data from year 1954 – 2013.

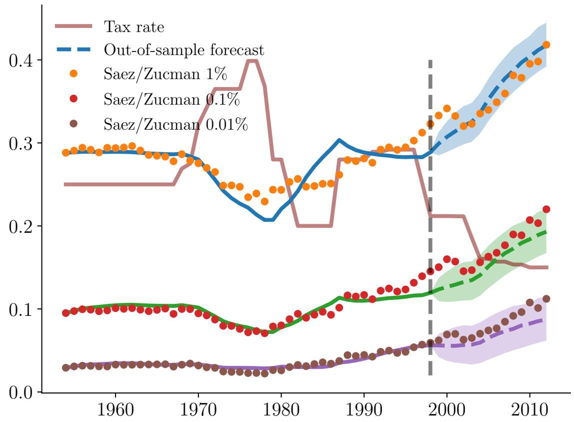

Figure 7: Out-of-sample testing using the time series from years 1954 – 1998 and the 3-dimensional Kalman Filter.
Notes: Filtered series until year 1998, from then on the dashed lines represent the median of simulations together with the 95% prediction intervals.

Although we add additional time series to the filter, we do not allow different series of shocks for each series of data. Hence, all series are subject to the same exogenous disturbances. In Figure 6 we present the fitted model, while Figure 7 displays out of sample forecasts, once again starting in the year 1998. It can be seen from the parameter estimates that the quality of fit as measured by σ_s is actually increasing with the number of dimensions, even though with each series 59 new data points are included, but only one more parameter. Again, the 3-D case matches the percentiles quite well, whereas the top 0.1 and 0.01% are characterized by an even better fit. Moreover, there are some further insights. Again the spike in inequality following the year 2000 can not be explained well by the change in taxes only, a result that holds for all three time series of percentiles.

[24] For a true log-normal distribution the estimated Pareto tail in the tails would diverge $\lim_{\hat{w}\to\infty} \hat{\alpha}(\hat{w}) = \infty$.

[25] The figure is generated using the time-varying solution of the Fokker-Planck equation, as presented in Appendix C.

While the decrease in relative wealth owned by the top 1% in the late 1970s and early 1980s is overpredicted by our model, it performs quite well for the top 0.1% and 0.01%. As above, the out-of-sample predictions suggest that the model explains the data reasonably well, implying its feasibility long-run forecasting.

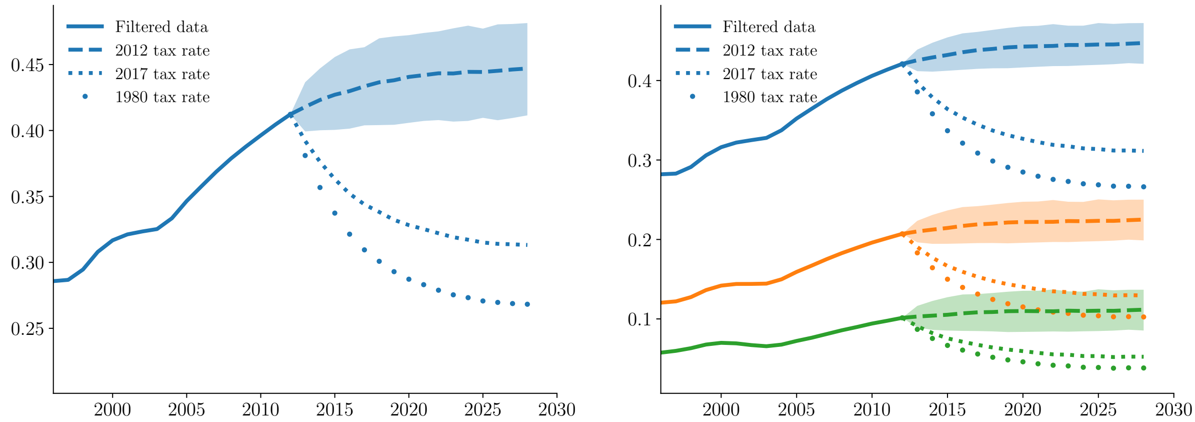

Figure 8: Projections taking the last observation in 2013 as a starting point and feeding in different tax series. Tax rates were at 15% in 2012, 23.8% in 2017 and 28% in 1980.

Notes: While the stationary level given the current tax rate is not reached before approximately 2025, moderate changes in the capital income tax have a significant effect on all top-shares. Left: one-dimensional Filter. Right: 3-dimensional Filter. Calibration as in Table 1.

In Figure 8 (left) we show forecasts given the using the estimated parameters and different tax regimes. As of 2017 the actual tax rate calculated as above is at 23.8%. This episode is not included in our series since Saez and Zucman (2016) only provide data until 2012. An unchanged tax regime would be sufficient to reverse the trend and bring inequality back to the level of the early 2000s. A further increase to 28%, which is the level from 1980 would lever the concentration back to its value from the 1990s. In this case the share of the top 1% would almost fall down to the level currently held by the top 0.1%, implying a considerable level of redistribution. On the other hand, tax cuts, as currently suggested, would result in a further increase in wealth inequality compared to the status quo. Since the data is not yet available we used the 2012-rate for these projections. The simulations also confirm our analytic result that the decrease in inequality after tax increases is faster than the ascent of inequality following tax reductions.

5 Conclusion

The main purpose of this work is to develop a simple, yet micro-founded portfolio selection model to explain the dynamics of wealth inequality given empirical tax series. Although a quite straightforward approach, this stands in contrast to the majority of the theoretical literature on wealth inequality which takes income inequality as a starting point.

We apply this model to the USA, which experienced a recent and substantial increase in wealth inequality. Due to the parsimonious nature of our model the degree of freedom to fit the empirical evidence is very limited. Nevertheless, our model matches the data surprisingly well, both in levels and also in transition speed. Our analytic results emphasize that the level and the transition speed of wealth inequality depend crucially on the degree of capital taxation, which is quantitatively and qualitatively in line with our results from the filtering process. We conclude that the given tax series have a very high explanatory power regarding the dynamics of US wealth distribution over the last 60 years.

This also implies that one answer on the policy question on how to influence the distribution of wealth – and potentially reverse the recent increase in wealth inequality observed in developed economies – can be given by looking at the tax system. An increase in capital gains taxes, or alternatively a gross tax on wealth as suggested in Piketty (2014), will very likely reduce wealth concentration and has the potential to upturn the observed trends. Our projections predict that, for the USA – continuing on the present path of capital taxation – the gap between rich and poor is expected to shrink whereas (*"massive"*) tax cuts will further increase the degree of wealth concentration.

There are two implications for future research. Although our model fits the data quite well, there are periods where it falls short of accounting for the data. First, we consider it important to identify whether the those shortcomings are due to bad tax data, measurement errors, or to reasons that are exogenous to our model. Second, if these reasons are exogenous it is crucial to investigate them further.

Further, the quality of the model's result severely hinges on the quality of the data. A better availability of data on wealth dispersion at higher frequencies would give better means to improve our model and enhance the understanding of the issue of wealth inequality in the 21st century.

References

Aiyagari, S. R. (1994). Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics 109*(3), 659–84.

Aiyagari, S. R. (1995). Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting. *Journal of Political Economy 103*(6), 1158–75.

Alstadsaeter, A., N. Johannesen, and G. Zucman (2017). *Tax Evasion and Inequality.* Working Paper.

Alvaredo, F., A. B. Atkinson, and S. Morelli (2017). Top Wealth Shares in the UK Over More than a Century. Dept. of Economics Research Paper Series No. 01/WP/2017, University Ca' Foscari of Venice.

Aoki, S. and M. Nirei (2017). Zipf's Law, Pareto's Law, and the Evolution of Top Incomes in the U.S. *American Economic Journal: Macroeconomics forthcoming.*

Benhabib, J., A. Bisin, and S. Zhu (2011). The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents. *Econometrica 79*(1), 123–157.

Benhabib, J., A. Bisin, and S. Zhu (2014). The distribution of wealth in the blanchard–yaari model. *Macroeconomic Dynamics FirstView*, 1–16.

Bewley, T. (1986). Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consumers. In W. Hildenbrand and A. Mas-Collel (Eds.), *Contributions to Mathematical Economics in Honor of Gerard Debreu.* Amsterdam: North-Holland.

Blanchet, T., J. Fournier, and T. Piketty (2017). Generalized Pareto Curves : Theory and Applications. Working Papers 201703, World Inequality Lab.

Boehl, G. (2017). *On the Evolutionary Fitness of Rational Expectations. In preparation.*

Branch, W. A. (2004). The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations*. *The Economic Journal 114*(497), 592–621.

Cagetti, M. and M. D. Nardi (2009). Estate Taxation, Entrepreneurship, and Wealth. *American Economic Review 99*(1), 85–111.

Campbell, J. Y. and L. M. Viceira (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Number 9780198296942 in OUP Catalogue. Oxford University Press.

Cao, D. and W. Luo (2017). Persistent Heterogeneous Returns and Top End Wealth Inequality. *Review of Economic Dynamics 26*, 301–326.

Castaneda, A., J. Diaz-Gimenez, and J.-V. Rios-Rull (2003). Accounting for the U.S. Earnings and Wealth Inequality. *Journal of Political Economy 111*(4), 818–857.

Chan, S., J. Chu, and S. Nadarajah (2017, 3). Is the wealth of the forbes 400 lists really pareto distributed? *Economics Letters 152*, 9–14.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). Power-law distributions in empirical data. *SIAM Review 51*(4), 661–703.

Clemens, C. and M. Heinemann (2015). Endogenous growth and wealth inequality under incomplete markets and idiosyncratic risk. *Journal of Macroeconomics 45*(C), 300–317.

Craig, C. C. (1936, 03). On the frequency function of xy. *Ann. Math. Statist. 7*(1), 1–15.

Domeij, D. and J. Heathcote (2004). On The Distributional Effects Of Reducing Capital Taxes. *International Economic Review 45*(2), 523–554.

Fernholz, R. and R. Fernholz (2014). Instability and concentration in the distribution of wealth. *Journal of Economic Dynamics and Control 44*(C), 251–269.

Fischer, T. (2017). Thomas Piketty and the rate of time preference. *Journal of Economic Dynamics and Control 77*(C), 111–133.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). The dynamics of inequality. *Econometrica 84*(6), 2071–2111.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2017). Accounting for wealth inequality dynamics: Methods, estimates and simulations for france (1800-2014). CEPR Discussion Papers 11848, C.E.P.R. Discussion Papers.

Greenwood, R. and A. Shleifer (2014). Expectations of Returns and Expected Returns. *Review of Financial Studies 27*(3), 714–746.

Hamilton, J. (1994). Time series analysis. Princeton University Press.

Hommes, C. (2013). *Behavioral rationality and heterogeneous expectations in complex economic systems*. Cambridge University Press.

Hubmer, J., P. Krusell, and J. Anthony A. Smith (2016). The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation. NBER Working Papers 23011, National Bureau of Economic Research, Inc.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control 17*(5-6), 953–969.

Judd, K. L. (1985). Redistributive taxation in a simple perfect foresight model. *Journal of Public Economics 28*(1), 59–83.

Kalman, R. E. et al. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering 82*(1), 35–45.

Kaplan, G., B. Moll, and G. L. Violante (2016). Monetary Policy According to HANK. NBER Working Papers 21897, National Bureau of Economic Research, Inc.

Karlin, S. and H. Taylor (1981). *A Second Course in Stochastic Processes*. Number 2. Academic Press.

Kasa, K. and X. Lei (2017). Risk, Uncertainty, and the Dynamics of Inequality. Discussion Papers dp17-06, Department of Economics, Simon Fraser University, forthcoming in *Journal of Monetary Economics*.

Kaymak, B. and M. Poschke (2016). The evolution of wealth inequality over half a century: The role of taxes, transfers and technology. *Journal of Monetary Economics 77*(C), 1–25.

Krusell, P. and T. Smith (1998). Income and Wealth Heterogeneity in the Macroeconomy. *Journal of Political Economy 106*(5), 867–896.

Kumhof, M., R. Rancière, and P. Winant (2015). Inequality, Leverage, and Crises. *American Economic Review 105*(3), 1217–1245.

Lansing, K. J. (1999). Optimal redistributive capital taxation in a neoclassical growth model. *Journal of Public Economics 73*(3), 423–453.

Levhari, D. and T. N. Srinivasan (1969). Optimal savings under uncertainty. *The Review of Economic Studies 36*(2), 153–163.

Lundberg, J. and D. Waldenström (2017). Wealth inequality in sweden: What can we learn from capitalized income tax data? *Review of Income and Wealth* (77(1)), 1–25.

Mankiw, N. G., R. Reis, and J. Wolfers (2003). Disagreement about inflation expectations. Technical report, National Bureau of Economic Research.

Pareto, V. (1896). *Cours d'Economie Politique*. Genève: Droz.

Pfajfar, D. and E. Santoro (2008). *Asymmetries in inflation expectation formation across demographic groups*. University of Cambridge, Faculty of Economics.

Pfajfar, D. and E. Santoro (2010). Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior & Organization 75*(3), 426–444.

Piketty, T. (2014). *Capital in the twenty-first century*. Boston: Harvard University Press.

Piketty, T. and G. Zucman (2014). Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics 129*(3), 1255–1310.

Pulley, L. B. (1983). Mean-variance approximations to expected logarithmic utility. *Operations Research 31*(4), 685–696.

Roine, J. and D. Waldenström (2015). Chapter 7 - long-run trends in the distribution of income and wealth. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume 2 of *Handbook of Income Distribution*, pp. 469 – 592. Elsevier.

Saez, E. and S. Stantcheva (2016). A Simpler Theory of Optimal Capital Taxation. NBER Working Papers 22664, National Bureau of Economic Research, Inc.

Saez, E. and G. Zucman (2016). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *The Quarterly Journal of Economics 131*(2), 519–578.

Singer, A., Z. Schuss, A. Osipov, and D. Holcman (2008). Partially reflected diffusion. *SIAM Journal on Applied Mathematics 68*(3), 844–868.

Stiglitz, J. E. (1969). The Effects of Income, Wealth, and Capital Gains Taxation on Risk-Taking. *The Quarterly Journal of Economics 83*(2), 263–283.

Straub, L. and I. Werning (2014). Positive Long Run Capital Taxation: Chamley-Judd Revisited. NBER Working Papers 20441, National Bureau of Economic Research, Inc.

Toda, A. A. (2012). The double power law in income distribution: Explanations and evidence. *Journal of Economic Behavior & Organization 84*(1), 364 – 381.

Vermeulen, P. (2016). Estimating the Top Tail of the Wealth Distribution. *American Economic Review 106*(5), 646–650.

Wold, H. O. A. and P. Whittle (1957). A model explaining the pareto distribution of wealth. *Econometrica 25*(4), 591–595.

Appendix

A Proof of Proposition 1

The product-normal distribution is treated extensively in Craig (1936). The probability distribution function is given by

$$f(z_{PN}) = \frac{1}{\pi} K_0 \left(|z_{PN}| \right), \tag{26}$$

with $z_{PN} \equiv \epsilon^1 \epsilon^2$ with $\epsilon^i \sim N(0,1)$ and K_0 being the modified Bessel-function of the second kind. The function is symmetric around the mean of zero and exhibits leptokurtic behavior. It is more appealing to write this using the Moment-Generating Function (MGF), which in this case is given by

$$M_{Z_{PM}}(t) = \frac{1}{\sqrt{1-t^2}}. \tag{27}$$

Using this it is easy to show that the mean and skewness are zero, while the standard deviation is given by

$$SD(z_{PN}) = 1. \tag{28}$$

This distribution is highly comparable to the Laplace distribution. For a zero-mean the probability density function of the latter is given by

$$f(z_L) = \frac{1}{2\lambda} \exp\left(-\frac{|z_L|}{\lambda} \right) \tag{29}$$

for shape parameter $\lambda > 0$, having both a mean and a skewness of zero. The standard deviation of Laplace is

$$SD(z_L) = \sqrt{2}\lambda. \tag{30}$$

The Laplace distribution is also very appealing as each half takes the form of an exponential function. The moment generating function of the Laplace distribution is

$$M_{Z_L}(t) = \frac{1}{1 - \lambda^2 t^2}. \tag{31}$$

Comparing this with the MGF of the product-normal distribution it becomes obvious that the two are not identical. In fact, the sum of two product-normal variables follows a Laplace distribution.[26]

As a reasonable approximation we replace the product-normal with the Laplace distribution. To obtain the shape parameter λ that best approximates the standard normal product distribution we equalize the second order Taylor expansions of both MGFs around $t = 0$, which in fact is equivalent to choosing λ to match the first two moments of the function. This yields

$$\begin{aligned}
\sum_{n=0}^{2} \frac{\partial^n M_{Z_{PM}}(0)}{n!\partial t^n}(t-0)^n &= \sum_{n=0}^{2} \frac{\partial^n M_{Z_L}(0)}{n!\partial t^n}(t-0)^n \\
1 + \frac{t^2}{2} &= 1 + t^2\lambda^2 \\
\lambda &= \sqrt{\frac{1}{2}} \approx 0.707. \quad \blacksquare
\end{aligned}$$

B Proof of Proposition 4

The solution to the Fokker-Planck equation can be easily determined using the Laplace transform into the frequency domain[27] given by

$$\mathcal{L}\{f(\hat{w}, t)\} \equiv F(s, t) \equiv \int_0^\infty f(\hat{w}, t) \exp(-s\hat{w}) d\hat{w}. \tag{32}$$

[26] Using the MGF it is easy to show that if there are four independently distributed normal shocks with zero mean $X_i \sim N(0, \sigma_i)$ and we have $\sigma_1\sigma_2 = \sigma_3\sigma_4$ then $X_1X_2 + X_3X_4$ follows a Laplace distribution with zero mean and $\lambda = 1$.

[27] This procedure is also employed in Gabaix et al. (2016) and Kasa and Lei (2017) to solve similar problems.

The latter is of particular help for solving linear differential equations as the n-th derivative is given by $\mathcal{L}\{f^n(\hat{w})\} = s^n F(s,t)$. For the right tail (index r) the characteristic equation is given by

$$\frac{\partial F(s,t)}{\partial t} = \mu s F(s,t) + 0.5\delta^2 s^2 F(s,t) = \Lambda_r(s)F(s,t) \tag{33}$$

with $\Lambda_r(s) = \mu s + 0.5\delta^2 s^2$. The stationary solution is found by setting $\frac{\partial F(s,t)}{\partial t} \overset{!}{=} 0$, leading to

$$\Lambda_r(s) = 0 \rightarrow s_r = -\frac{2\mu}{\delta^2} \equiv -\alpha. \tag{34}$$

In this case, the cross-sectional distribution of log wealth $\hat{w} \equiv \ln(w)$ is given by an exponential distribution, while wealth follows a Pareto distribution. The value α is the rate parameter of the exponential distribution respectively the Pareto coefficient.

This approach can also be employed to make a statement about the convergence rate. As our paper only considers the top shares we focus on the right tail of the distribution, as described by $\Lambda_r(s)$. In fact the convergence rate of the n-th moment $E(\hat{w}^n)$ is given by $\Lambda_r(-n)$. For the example of the mean it would be

$$\Lambda_r(-1) = -\mu + 0.5\delta^2 = -\frac{\tau}{\lambda}. \tag{35}$$

It is well known that for the Pareto distribution only moments with $0 < n < \alpha$ exist. For the parametrization to fit the US wealth distribution we always have $\alpha < 2$. The average convergence time - as defined in Gabaix et al. (2016) - emerges for $\bar{n} = 0.5\alpha = \frac{\mu}{\delta^2}$. It is given by

$$\Lambda_r\left(s = -\frac{\mu}{\delta^2}\right) = -0.5\frac{\mu^2}{\delta^2} < 0 \tag{36}$$

Assume that the distribution starts at a stationary distribution $F(s,0)$. After a shock in parameters the new stationary distribution is $F(s,\infty)$. Solving the differential equation 33, we find the convergence in the frequency domain for some s is given by

$$F(s,t) = F(s,\infty) + [F(s,0) - F(s,\infty)]\exp(\Lambda_r(s)t). \tag{37}$$

In this case, we have $\Lambda_r(s) = -\phi = -\frac{\mu^2}{2\delta^2}$ as the average convergence rate. More generally we can write it as

$$F(s,t+\tau) = F(s,\infty) + [F(s,t) - F(s,\infty)]\exp(\Lambda_r(s)\tau), \tag{38}$$

which for our special case of $\tau = 1$ implies

$$F(s, t+1) = F(s, \infty) + [F(s,t) - F(s, \infty)] \exp(\Lambda_r(s)). \quad (39)$$

C Details on the time-dependent cross-sectional distribution

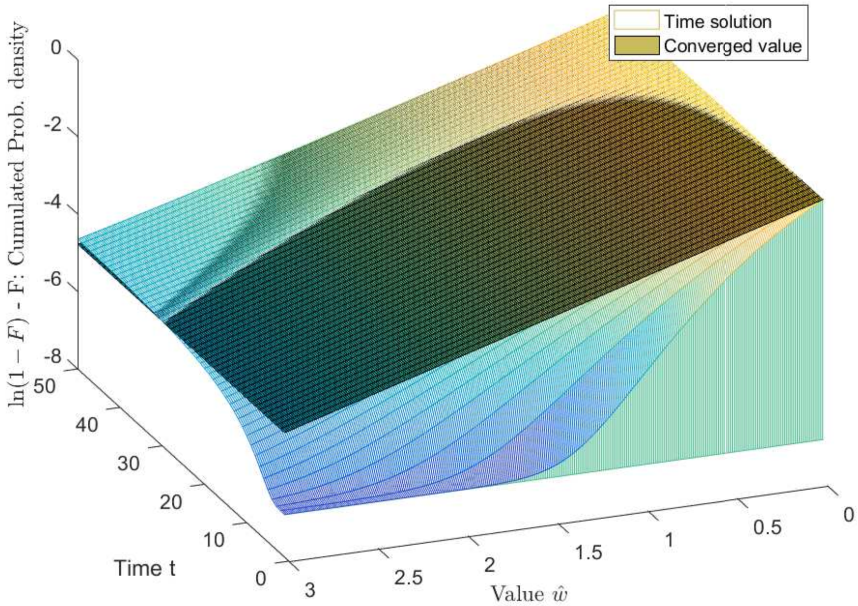

Figure 9: Log-Cumulative probability density function: Stationary solution and time-dependent solution

(Singer et al., 2008, p. 853) provide a full solution of the underlying Fokker-Planck equation in the time domain for some given initial value $\hat{w}_0 = 0$. We also want to assume that the reflecting boundary is $\hat{w}_{min} = 0$ (i.e. $w_{min} = 1$). The solution is given by

$$f(\hat{w}, t) = \frac{1}{\sqrt{\pi t 2\delta^2}} \exp\left(-\frac{\hat{w}^2}{2\delta^2 t}\right) \exp\left(-0.5\alpha\hat{w} - \frac{1}{8}\alpha^2\delta^2 t\right) + f(\hat{w}, \infty)\Phi\left(-\frac{\hat{w}}{\delta\sqrt{t}} + 0.5\alpha\delta\sqrt{t}\right), \quad (40)$$

for which $f(\hat{w}, \infty) = C \exp(-\alpha\hat{w})$ describes the long-run stationary solution and Φ is the cumulative probability density function of the normal distribution. We have $C = \alpha w_{min}^{\alpha} = \alpha$. For small time values t it is Gaussian, finally converging to an exponential distribution. In terms of transformed values $w = \exp(\hat{w})$ this implies a transformation from log-normal to Pareto.

It is evident that the solution is both a function of time t and the value of $\hat{w}$. Essentially, the function slowly *fattens out* to the tails (cf. figure 9). Thus, the measured Pareto tail $\hat{\alpha}$ decrease in time, but increases with the value of $\hat{w}$. Technically, it never converges in the fattest tails ($\lim_{\hat{w}\to\infty} f(\hat{w}, t \to \infty) \neq f(\hat{w}, \infty)$).

Acknowledging that the first part is a normal distribution with zero mean and variance $\delta^2 t$ (exploding in time) and abbreviating this with $f_0(\hat{w}, t)$ as well as using the definition of the average convergence speed $\phi = \frac{\mu^2}{2\delta^2} = \frac{1}{8}\alpha^2\delta^2$, we can write:

$$f(\hat{w}, t) = f_0(\hat{w}, t) \exp(-0.5\alpha\hat{w}) \exp(-\phi t) + f(\hat{w}, \infty)\Phi\left(-\frac{\hat{w}}{\delta\sqrt{t}} + 0.5\alpha\delta\sqrt{t}\right). \qquad (41)$$

The very last term in the equation related to the normal CPDF captures both the convergence speed $(1 - \exp(-\phi t))$ and the non-linearity adjustment for $f(\hat{w}, \infty)$. It is obvious that this is incorporated in a non-trivial manner. In the empirical application we choose a simplified non-linearity adjustment not least to keep the estimation feasible.

ONLINE APPENDIX NOT INTENDED FOR PUBLICATION

D Details on Kalman filter and estimation procedure

The most general form of the Kalman filter applies for a Linear Gaussian model specified by

$$x_{t+1} = A_t x_t + b_t + \mathcal{N}(0, Q_t), \tag{42}$$

$$z_t = C_t x_t + d_t + \mathcal{N}(0, R_t), \tag{43}$$

where $\{z_t\}_0^T$ is the series of the vector of observables and $\{x_t\}_0^T$ is the series of the vectors of unobservable true variables. A_t is the state transition matrix between times t and $t+1$ at time t and b_t are the state offsets. C_t is the observation matrix and d_t the observation offset, which are here assumed to be the identity matrix or the zero vector respectively. Q_t is the state transition covariance matrix, which is assumed to be filled with σ_s^2, since we assume that the shock series for each dimension is perfectly correlated. Finally, R_t is the observation covariance matrix that is filled by the variance of the assumed measurement error, that we will here denote by σ_m^2. Taking the two-dimensional model as an example, the above can be rewritten as

$$s_{t+1} = \text{diag}(\rho_t) s_t + (1-\rho_t) \begin{bmatrix} s^{1\%}(\tau_t, \infty) \\ s^{0.1\%}(\tau_t, \infty) \end{bmatrix} + \mathcal{N}\left(0, \begin{bmatrix} \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 \end{bmatrix}\right), \tag{44}$$

$$\hat{s}_t = s_t + \mathcal{N}\left(0, \begin{bmatrix} \sigma_m^2 & \sigma_m^2 \\ \sigma_m^2 & \sigma_m^2 \end{bmatrix}\right), \tag{45}$$

where s_t is the vector of top 1% and top 0.1%-shares that is assumed to be the *true* at time t, and $\hat{s}_t$ is the vector containing the empirical data/top shares at each point in time.

The (real-time) Kalman Filter is an algorithm designed to estimate $P(x_t|z_{0:t})$, i.e. the probability of x_t at time t given the series of observations until then. The exact mathematical representation of the Kalman procedure would be too lengthy here and is relatively standard in time series analysis. We therefore redirect to standard works such as for instance to (Hamilton, 1994, Chapter 13). Similarly, the (post-processing) Kalman Smoother is an algorithm designed to estimate $P(x_t|z_{0:T-1})$. Let us define $\Gamma = (\gamma_{1\%}, \gamma_{0.1\%}, \gamma_{0.01\%,}, \sigma_s)$ to be the set of parameters we seek to estimate. For our estimation we aim to find

$$\max_{\Gamma} P(s_{0:T-1}|\hat{s}_{0:T-1};\Gamma). \tag{46}$$

Let us define $L(s_{0:T-1},\Gamma) = \log P(s_{0:T-1}|\hat{s}_{0:T-1};\Gamma)$ and estimate Γ by numerically maximizing

$$\arg\max_{\Gamma} \mathbb{E}_{s_{0:T-1}}[L(s_{0:T-1},\Gamma)|\hat{s}_{0:T-1},\Gamma]. \tag{47}$$

The filtered series then represent $s_{0:T-1}$ as in

$$P(s_{0:T-1}|\hat{s}_{0:T-1},\Gamma). \tag{48}$$

In this work we make use of the pykalman package and use standard numerical maximzation procedures provided by numpy for the platform Python.[28]

E Robustness of the estimates with respect to the sample period

In Figure 10 the estimates of the two central parameters using the one-dimensional filter are plotted depending on the sample used for the estimation process. Leaving out the last 12 periods (top figure) does not have a significant impact on the estimates. However, excluding too many of the recent periods which display exceptionally high inequality and fast transition dynamics does decrease the estimate of $\hat{\gamma}$. This is well in line with our comparative statics exercise in Section 3.4. In parallel the standard deviation of exogenous noise $\hat{\sigma}_s$ decreases when excluding more periods at the end. Thus, these last episodes increase $\hat{\gamma}$ to an extent that overall more exogenous variation $\hat{\sigma}_s$ is necessary to explain the data. From a quantitative perspective these effects are, however, marginal.

More crucial for our results are the initial periods (bottom figure). Since this period is characterized by a relatively stable degree of wealth-concentration the value of $\hat{\gamma}$ increases quickly when ignoring the years at the beginning for the estimation. Moreover, the level of $\hat{\sigma}_s$ increases, indicating a looser fit that requires more exogenous variation. Overall, both results suggest that γ is probably increasing over time.

[28]The packages and documentations can be found at https://pykalman.github.io/ and http://www.numpy.org/.

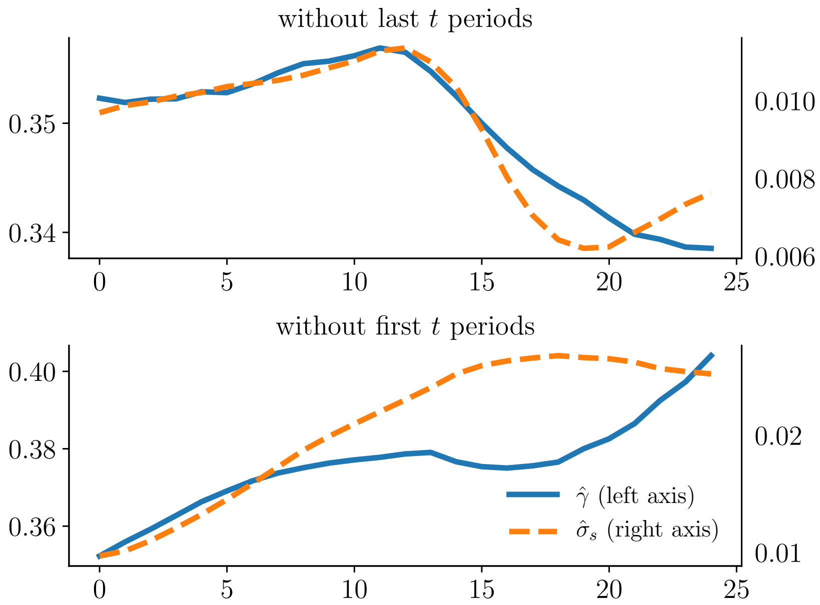

Figure 10: Changes in the estimates of $\hat{\gamma}$ and $\hat{\sigma}_s$ when using shorter data samples.

F Two-Dimensional Kalman filter

For completeness we provide the series of the two-dimensional Kalman filter in Figure 11. These are well in line with the simulations from the main body and merely provided for completeness.

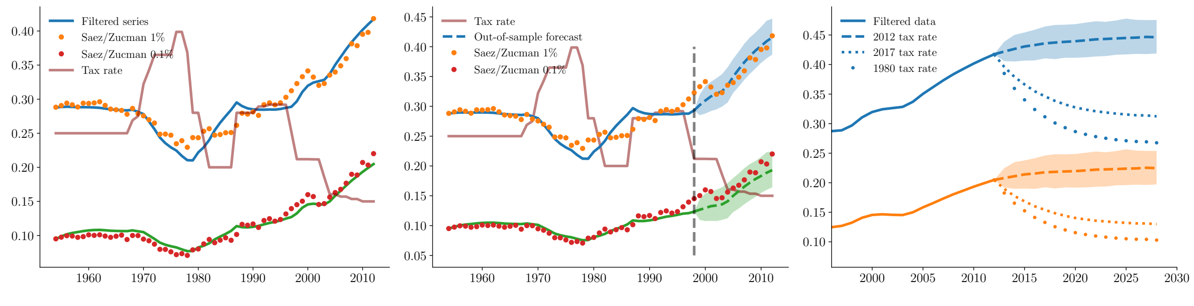

Figure 11: Left: Kalman filter using the the series of the top 1% and top 0.1%. Middle: Out-of-sample forecasts. Right: Projections using different tax series. Estimated values are provided in Table 1.

IMFS Working Paper Series

Recent Issues

117 / 2017	Tobias H. Tröger	Why MREL Won't Help Much
116 / 2017	Tobias H. Tröger	Too Complex to Work – A Critical Assessment of the Bail-in Tool under the European Bank Recovery and Resolution Regime
115 / 2017	Guenter W. Beck Volker Wieland	How to Normalize Monetary Policy in the Euro Area
114 / 2017	Michael Binder Jorge Quintana Philipp Lieberknecht Volker Wieland	Model Uncertainty in Macroeconomics: On the Implications of Financial Frictions
113 / 2017	Mewael F. Tesfaselassie Maik Wolters	The Impact of Growth on Unemployment in a Low vs. a High Inflation Environment
112 / 2017	Gerhard Rösl Franz Seitz Karl-Heinz Tödter	Doing away with cash? The welfare costs of abolishing cash
111 / 2017	Jinhyuk Yoo	Capital Injection to Banks versus Debt Relief to Households
110 / 2017	Robert C. M. Beyer Volker Wieland	Instability, imprecision and inconsistent use of equilibrium real interest rate estimates
109 / 2016	Helmut Siekmann	Replacing or Supplementing the Euro in Member States whose Currency is the Euro
108 / 2016	Helmut Siekmann	Restricting the Use of Cash in the European Monetary Union
107 / 2016	Volker Wieland Elena Afanasyeva Meguy Kuete Jinhyuk Yoo	New Methods for Macro-Financial Model Comparison and Policy Analysis
106 / 2016	Helmut Siekmann	Konstruktionsfehler bei der Einlagensicherung auf EU-Ebene
105 / 2016	Athanasios Orphanides	Fiscal Implications of Central Bank Balance Sheet Policies
104 / 2016	Helmut Siekmann	Preis- und Finanzstabilität: der Primat der Politik, der rechtliche Rahmen und das „ökonomische Gesetz“

103 / 2016	John B. Taylor Volker Wieland	Finding the Equilibrium Real Interest Rate in a Fog of Policy Deviations
102 / 2016	Tilman Bletzinger Volker Wieland	Forward guidance and “lower for longer”: The case of the ECB
101 / 2016	Helmut Siekmann	Ziele, Aufgaben und Befugnisse des Europäischen Systems der Zentralbanken (ESZB) (publ. in: Scherzberg / Can / Doğan (eds.), Die Sicherung der Geldwert- und Finanzstabilität in Deutschland und in der Türkei, 2016, pp. 79-118)
100 / 2016	Robert C. M. Beyer Volker Wieland	Schätzung des mittelfristigen Gleich-gewichtszinses in den Vereinigten Staaten, Deutschland und dem Euro-Raum mit der Laubach-Williams-Methode
99 / 2015	Helmut Siekmann	Exit, Exclusion, and Parallel Currencies in the Euro Area
98 / 2015	Helmut Siekmann	Stellungnahme für die öffentliche Anhörung der Verfassungskommission des Landtags Nordrhein-Westfalen zur Schuldenbremse
97 / 2015	José María Liberti Amit Seru Vikrant Vig	Information, Credit, and Organization
96 / 2015	Patrick Tuschl	Rechtsgrundlagen für die Vereinbarkeit von Finanzhilfen für Kreditinstitute mit dem Beihilferecht der EU im Rahmen der Finanzmarktkrise
95 / 2015	Athanasios Orphanides	Fear of Liftoff: Uncertainty, Rules and Discretion in Monetary Policy Normalization
94 / 2015	Bettina Brüggemann Jinhyuk Yoo	Aggregate and Distributional Effects of Increasing Taxes on Top Income Earners
93 / 2015	Tobias H. Tröger	Regulatory Influence on Market Conditions in the Banking Union
92 / 2015	Janis Skrastins Vikrant Vig	How Organizational Hierarchy Affects Information Production
91 / 2015	Florian Hoffmann Roman Inderst Marcus Opp	Regulating Deferred Incentive Pay
90 / 2015	Helmut Siekmann	The Legality of Outright Monetary Transactions (OMT) of the European System of Central Banks (publ. in: Rövekamp / Bälz / Hilpert (eds.), Central Banking and Financial Stability in East Asia, 2015, pp. 101-123)